DOCUMENT RESUME

ED 093 942 TM 003 760

AUTHOR Gupta, Ram K.; And Others
TITLE Establishing Construct Validity of Psychological Instruments Thru Item-Wise Concurrent Validation.
SPONS AGENCY Public Health Service (DHEW), Arlington, Va.
PUB DATE Apr 74
NOTE 7p.; Paper presented at the Annual Meeting of the American Educational Research Association (59th, Chicago, Illinois, April 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS Hypothesis Testing; Internal Scaling; *Measurement Techniques; Personality Tests; *Psychological Tests; *Test Construction; *Test Validity

ABSTRACT

A logically sound and methodologically easy technique for improving the construct validity of instruments and developed for relatively new concepts is described. It is based on the premise that if an instrument is valid, so should be each item within it and, therefore, the latter should sustain in field setting the behavioral implications derived from the theory underlying the construct. This idea was examined for 87 items of seven factorially "pure" scales, using a nearly random sample of about 7,000 subjects, and testing 10 such hypotheses for each item. The results are described. (Author/BB)

ED 093942

TM 003 760

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ESTABLISHING CONSTRUCT VALIDITY OF PSYCHOLOGICAL INSTRUMENTS THRU ITEM-WISE CONCURRENT VALIDATION[1]

Ram K. Gupta, Department of Educational Psychology, University of Alberta, Edmonton, Canada

W. Bruce Handley, Kelsea School Division, the Pas, Manitoba, Canada

Merton P. Strommen, Youth Research Center, Minneapolis, Minn.

(a) Objective

Developers of psychological tests and devices generally limit their attention to what Loevinger (1957) called the "substantive component" of construct validity and/or to the "structural component". These are analogous to "content" and "factorial" validities, respectively. Few seem to attend to her "external component" - that needing empirical substantiation of the construct, probably because a well-established criterion measure for the construct purported to be embodied in the newly developed instrument is rarely available. The present research is an attempt to remove this gap by making logically justifiable "apriori" behavioral predictions (using the theory underlying the construct) and then examining the same through hypothesis testing, using concurrently obtained data. It is hoped that the procedure will be accepted as being simple enough for adoption to refine newly developed instruments and to establish their construct validity.

(b) Theoretical Framework

Let us assume that a researcher has secured content validity (semantic adequacy of the title of an instrument) through (1) formulating a sound rationale for a particular concept or construct; (2) operationalizing the rationale

---

1. Paper presented to the annual meeting of the American Educational Research Association, in April 1974 at Chicago. This research was supported by Grant No. MH 16460-2 of the Department of Health, Education, and Welfare, Public Health Service. Computing facilities were kindly provided by the University of Alberta, Canada.

by reducing it to behavioral statements (or items), and (3) arriving at a consensus of experts' judgments on the statements. Let us further assume that he has also secured functional unity, cohesiveness, homogeneity or interal consistency (Cronbach, 1951; Hoyt, 1941; Kuder and Richardson, 1937; Loevinger, 1947, 1948) through results from suitable multivariate analyses (Gupta, 1968; Gupta and Burnett, 1972; Strommen and Gupta, 1971). The question now arises: does the instrument possess criterion related validity? A "yes" answer will instill confidence in the users and improve the interpretability of the scores. One could attempt an answer to the above question by correlating scores on the new instrument with those on another tried and proven instrument, measuring the same trait. Most often, however, the new instrument is probably the only one related to the construct under examination. The researcher has, therefore, to use an indirect approach (APA, 1954, p. 14).

A review of literature failed to provide guidance in regard to available indirect approaches. The researchers, therefore, used the following approach:

IF A SCALE DOES REALLY REFLECT THE CONSTRUCT FOR WHICH IT WAS DEVELOPED OR ABOUT WHICH THERE IS A HIGH DEGREE OF CONSENSUS AMONG EXPERTS, AND WHICH HAS HIGH INTERNAL CONSISTENCY, THEN IT SHOULD BE POSSIBLE TO DEMONSTRATE EMPIRICALLY THAT THE LOGICALLY JUSTIFIABLE, BEHAVIORAL PREDICTIONS MADE ON THE BASIS OF THE CONSTRUCT AND THE THEORY UNDERLYING IT DO, IN FACT, COME TRUE

The predicted behaviors may be actual, observable actions. Alternatively, they may be verbal responses to appropriate stimuli - questions, items, statements, etc. In either case, however, our knowledge about them has to come from sources completely independent of the scale itself for meaningful and convincing validation of the construct.

Work on the above lines yielded highly promising results on a high school population as shown below:

(c) Procedure

Seven scales were developed by Handley (1973) around the general theme of mutuality which, as interpreted here, implies a universal need typifying youths of all cultures and sub-cultures, a need to feel accepted by an identifiable family or group of persons. Through the process of acceptance, he comes to sense the feelings of warmth and belongingness, the reality of being important to others.

The scales, with their sizes and internal consistencies given within parentheses, were:

1. Peer Relations - General: concern over acceptance by others, specially by peers and teachers. ($n = 17$; $r_{11} = .875$)

2. Peer Relations - Opposite Sex. ($n = 12$, $r_{11} = .839$)

3. Self Confidence. ($n = 16$; $r_{11} = .862$)

4. School Relations and Problems: Concern over school grades, inability to study and concentrate. ($n = 8$; $r_{11} = .846$)

5. Personal Worth: Concern over not having accomplished much or not having lived up to one's own ideals. ($n = 14$; $r_{11} = .844$)

6. Family Unity: Concern over emotional climate within the home. ($n = 7$; $r_{11} = .862$)

7. Parental Understanding: Concern over a lack of communication and understanding with parents. ($n = 13$; $r_{11} = .899$)

The content validity of the scales was secured through experts' judgments and internal consistency through the use of factor and component analyses (Harris, 1964; Hotelling, 1933; Kaiser, 1956, 1958, 1960, 1964; Kaiser and Cafferey, 1965) and also through cluster analyses (DuBois, et al 1952, Gupta and Burnett, 1972; Loevinger, 1947, 1948), followed by differential response weighting through the use of the Method of Reciprocal Averages (Loevinger, et al 1953; Mosier, 1946).

For examining the criterion-related validity, it was hypothesized that the subjects who lack mutuality will be anxious about their relationships with others and show this through expressions of concern and, therefore, higher scores. Contrariwise, those who experience the warmth and acceptance of a familial group will be less troubled over their relationships with others and, therefore, will score lower, and that the higher and lower scores of the contrasting (or criterion) groups will be found on every item within each scale.

Contrasting groups were formed on each of ten selected items, none of which was included in any of the seven scales and each of which, therefore, provided independent information. They were:

1. I tend to be a lonely person (1. yes 2. no, sometimes)
2. I find life exciting and full of fun. (1. no, sometimes; 2. yes)
3. My parents are too strict. (1. very much; 2. somewhat; 3. never)
4. I have considered suicide (1. yes; 2.no)
5. How many close friends do you have? (1. fewer than 3; 2. three to seven; 3. eight or more)
6. To what degree do you feel trusted by your parents (or guardian)? (1. not at all; 2. somewhat; 3. very much)
7. Choose the number that best illustrates where you are in your circle of closest friends. (1. at the periphery; 2. in between the periphery and the center; 3. at the center)
8. I have trouble getting along with my father. (1. yes; 2. no)
9. I have trouble getting along with my mother. (1. yes; 2. no)
10. We have had serious difficulties in our home during the past year. (1. yes; 2. no)

On items 1, 2, 4, 8, 9, and 10, there were only two such groups on each. On the remaining four items, viz., 3, 5, 6, and 7, there were three groups on each. The difference between the means of the responses of the subjects were

examined for statistical significant differences through ANOVA. The total number of subjects involved was 7050, who constituted a nearly random sample of the ecumenical population of the USA in 1970.

(d) Results and Conclusions

The following rules of thumb were established about the items:

(1) If nine or all the ten F-ratios for an item were significant (alpha = .05), the item was regarded as having sustained its claim for membership in the scale.

(2) If such F-ratios were 7 or 8 (out of 10), the item was regarded as being reasonably good for the scale but needed to be kept under vigilance in follow-up studies.

(3) If such F-ratios were fewer than seven, the item was considered to have dubious claim for retention in the scale.

The above three rules yielded the following summary results:

Scale	Number of Items Satisfying		
	Rule 1	Rule 2	Rule 3
1. Peer Relations - General	17	0	0
2. Peer Relations - Opposite Sex	10	1	1
3. Self Confidence	16	0	0
4. School Relations and Problems	6	1	1
5. Personal Worth	12	1	1
6. Family Unity	7	0	0
7. Parental Understanding	13	0	0
TOTAL	81	3	3

It is concluded that the construct validity will improve if the items falling under rule 3 and also possibly those under rule 2 were excluded from their respective scales.

(e) Educational or Scientific Importance of the Study

The scientific importance of this research lies in pointing out to a (practically speaking) very simple, logically very natural and sound technique for improving the construct validity of newly developed instruments.

BIBLIOGRAPHY

American Psychological Association, American Educational Research Association, and National Council on Measurements Used in Education. "Technical Recommendations for Psychological Tests and Diagnostic Techniques, "Psychological Bulletin, 1954, 51 (Suppl.).

Cronbach, L.J., Coefficient alpha and the internal structure of tests, Psychometrika, 1951, 16, 297-334.

DuBois, P.H., Gleser, G.C. and Loevinger, J. The Construction of Homogeneous Keys for a Biographical Inventory. Res. Bull. 52-18, Air Training Command, Human Resources Research Centre, 1952, Lackland Air Force Base.

Gupta, R.K. "Certain Techniques of Multivariate Analysis Applied to Different Measures of Inter-Item Relationships for Developing Unifactor Tests," Journal of Educational Measurement, 1968, 5, 223-230.

Gupta, R.K. and Burnett, J.D. "A Program to Carry Out Cluster Analysis by Homogenous Grouping," Educational and Psychological Measurement, 1972, 32, 185-190.

Handley, Bruce, Substantive structural and external components of construct validity in instrumentation, Unpublished Ph.D. thesis, University of Alberta, 1973.

Harris, C.W., Some recent developments in factor analysis, Edu. Psy. Meast., 1964 , 24, 193-206.

Hotelling, H., Analysis of a complex of statistical variables into principal components. J. of Edu. Psy., 1933, 24, 498-520.

Hoyt, C.J. "Test Reliability Estimated by Analysis of Variance", Psychometrika, 1941, 6, 153-160.

Kaiser, F., The varimax criterion for analytic rotation in factor analysis, Psychometrika, 1958, 23, 187-200.

Kaiser, F., The application of electronic computers to factor analysis, Edu. Psy. Meast., 1960, 20, 141-51.

Kaiser, F. Psychometric approaches to factor analysis, Proceedings of the 1964 Invitational Conference on Testing Problems, Princeton, Educational Testing Service, 1964, 37-45.

Kaiser, F. and Caffrey, J., Alpha factor analysis, Psychometrika, 1965, 30, 1-14.

Kuder, G.F. and Richardson, M.W. "The Theory of the Estimation of Test Reliability," Psychometrika, 1937, 2, 151-160.

Loevinger, J. "A Systematic Approach to the Construction and Evaluation of Tests of Ability," Psychological Monographs, 1947, 61, 49pp.

Loevinger, J. "The Technic of Homogeneous Tests Compared with Some Aspects of Scale Analysis and Factor Analysis," Psychol. Bull., 1948, 45, 507-530.

Loevinger, J. "Objective Tests as Instruments of Psychological Theory," Psychological Reports, 1957, 3, 635-694.

Loevinger, J., Gleser, B., and DuBois, P. Maximizing the discriminating power of a multi-score test. Psychometrika, 1953, 18, 309-317.

Mosier, Charles I. "On the Reliability of a Weighted Composite," Psychometrika, 1943, 8, 161-168.

Mosier, C. "Machine Methods of Scaling by Reciprocal Averages," Proceedings of Research Forum. International Business Machine Corporation, 1946.

Strommen, M.P. and Gupta, R.K. Manual for Youth Research Survey Section 4. Minneapolis: Youth Research Centre, 1971.